IN THE CLAIMS:

Amend the claims as follows:

1-11 (Cancelled)

12. (Currently amended) A computer-implemented method of detecting duplicate documents in a network crawling system, comprising, at a server having one or more processors and memory:

constructing a plurality of tables, each table corresponding to a portion of a document address space, storing information identifying documents having a same document content identifier and each identified document having an associated document rank; wherein documents having the same document content identifier have the same content and documents having different document content identifiers have different content;

receiving a newly crawled document, such document characterized by a document content identifier and a document rank;

reading information stored in the plurality of tables to identify a set of documents sharing the document content identifier of the newly crawled document, and ascertaining an original representative document for the identified set of documents;

updating the information stored in at least one of the tables in accordance with the document ranks of the identified set of documents and the newly crawled document;

determining a representative document for the newly crawled document and the identified set of documents;

indexing the representative document when the representative document is the newly crawled document; and

repeating the receiving, reading, updating, determining and indexing operations with respect to a plurality of newly crawled documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the newly crawled documents are determined to be representative documents and are indexed.

13. (Previously presented) The method of claim 12, wherein information identifying the identified set of documents, including a particular document serving as the original representative document of the identified set, is stored in one or more tables.

14. (Previously presented) The method of claim 13, wherein the determining includes

comparing the document rank of the newly crawled document with that of the particular document from the identified set in accordance with a set of predefined comparison criteria;

selecting the newly crawled document as the representative document if the set of predefined comparison criteria are met; and

keeping the particular document as the representative document if the set of predefined comparison criteria is not met.

15. (Previously presented) The method of claim 14, wherein the set of predefined comparison criteria comprise at least two parameters, one parameter for comparison with an absolute difference of document ranks between the newly crawled document and the particular document, and another parameter for comparison with a ratio of document ranks between the newly crawled document and the particular document.

16. (Original) The method of claim 12, wherein the updating includes inserting information identifying the newly crawled document into the at least one table only when a predefined insertion condition is satisfied.

17. (Previously presented) The method of claim 16, wherein the predefined insertion condition is that the document rank of the newly crawled document is higher than the document rank of at least one document in the identified set of documents.

18. (Currently amended) A computer-implemented method of detecting duplicate documents in a network crawling system, comprising, at a server having one or more processors and memory:

constructing a plurality of tables, each table corresponding to a segment of a document address space, storing information identifying documents having a same document content identifier and each identified document having an associated document rank, wherein the plurality of tables comprise N+1 tables where N is an integer greater than one, wherein the N+1 tables comprise N tables, each generated during a respective phase of a set of N crawling phases, and a current table generated during a current one of the N crawling phases, wherein an oldest one of the N tables was generated during a previous instance of the current crawling phase;

receiving a newly crawled document, such document characterized by a document content identifier and a document rank; wherein documents having the same document

content identifier have the same content and documents having different document content identifiers have different content;

reading information stored in the N+1 tables to identify a set of documents sharing the document content identifier of the newly crawled document, and ascertaining an original representative document for the identified set of documents;

updating the information stored in the current table in accordance with the document rankings of the identified set of documents and the newly crawled document;

determining a representative document for the newly crawled document and the identified set of documents;

indexing the representative document when said representative document is the newly crawled document;

repeating the receiving, reading, updating, determining and indexing operations with respect to a plurality of newly crawled documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the newly crawled documents are determined to be representative documents and are indexed; and

upon completion of the current crawling phase, retiring the oldest one of the N tables.

19. (Original) The method of claim 18, wherein the reading comprises reading from a merged table that stores information from a plurality of the N tables, and reading from the current table.

20. (Previously presented) The method of claim 18, wherein information identifying the identified set of documents, including a particular document serving as the original representative document of the identified set, is stored in one or more tables.

21-36. (Cancelled)

37. (Currently amended) A system for detecting duplicate documents during network crawling, comprising:

one or more central processing units for executing programs;

a network interface for receiving documents; and

a duplicate document detection engine executable by the one or more central processing units, the engine comprising:

a plurality of tables, each table corresponding to a segment of a document address space, storing information identifying documents having a same document content identifier and each identified document having an associated document rank, wherein the plurality of tables comprise N+1 tables where N is an integer greater than one, wherein the N+1 tables comprise N tables, each generated during a respective phase of a set of N crawling phases, and a current table generated during a current one of the N crawling phases, wherein an oldest one of the N tables was generated during a previous instance of the current crawling phase;

instructions for receiving a newly crawled document, such document characterized by a document content identifier and a document rank; wherein documents having the same document content identifier have the same content and documents having different document content identifiers have different content;

instructions for reading information stored in the N+1 tables to identify a set of documents, sharing the document content identifier of the newly crawled document, and ascertaining an original representative document for the identified set of documents;

instructions for updating the information stored in the current table in accordance with the document rankings of the identified set of documents and the newly crawled document;

instructions for determining a representative document for the newly crawled document and the identified set of documents;

instructions for indexing the representative document when said representative document is the newly crawled document;

instructions for repeating the receiving, reading, updating, determining and indexing operations with respect to a plurality of newly crawled documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the newly crawled documents are determined to be representative documents and are indexed; and

instructions for retiring the oldest one of the N tables upon completion of the current crawling phase.

38. (Original) The system of claim 37 wherein the reading comprises reading from a merged table that stores information from a plurality of the N tables, and reading from the current table.

39. (Previously presented) The system of claim 37, wherein the identified set of documents, including a particular document serving as the original representative document of the identified set, are stored in one or more tables.

40. (Currently amended) A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for constructing a plurality of data structures for storing information of documents, each document characterized by a document content identifier and a document rank, the information stored in the plurality of data structures include the document content identifier and a document rank for each document; wherein documents having the same document content identifier have the same content and documents having different document content identifiers have different content;

instructions for receiving a requesting document in association with its document content identifier and document rank;

instructions for selecting from the plurality of data structures a set of documents sharing the same document content identifier as the requesting document, and ascertaining an original representative document for the identified set of documents;

instructions for generating a new set of documents from the requesting document and the selected set of documents in accordance with their document rank;

instructions for identifying a representative document of the new set of documents;

instructions for indexing the representative document when said representative document is the requesting document; and

instructions for repeating the receiving, selecting, generating, identifying, and indexing operations with respect to a plurality of requesting documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the requesting documents are determined to be representative documents and are indexed.

41. (Canceled)

42. (Original) The computer program product of claim 40, wherein the plurality of data structures include a data structure for storing information of multiple sets of documents, each set of documents sharing a same document content.

43. (Original) The computer program product of claim 40, wherein the plurality of data structures include a data structure for storing information of multiple sets of documents, each set of documents sharing a same document address.

44. (Currently amended) The computer program product of claim 40, wherein the document content identifier is a fixed length fingerprint of document content of a document characterized by the document content identifier.

45. (Currently amended) The computer program product of claim 40, wherein the document content identifier is a fixed length fingerprint of an address of a document characterized by the document content identifier.

46. (Previously presented) The computer program product of claim 40, wherein the generating instructions include

sorting the requesting document and the selected set of documents in accordance with a metric included in score information of the requesting document and selected set of documents; and

selecting a new set of documents, having at most a predefined number of documents, from the requesting document and the selected set of documents based on the sorting result.

47. (Original) The computer program product of claim 40, wherein

the score information for each document includes a document rank; and

the identifying instructions include

comparing the document rank of the requesting document with that of a particular document from the selected set of documents in accordance with a set of predefined comparison criteria, wherein the particular document was previously determined to be the representative document for the selected set of documents;

selecting the requesting document as the representative document for the new set of documents if the set of predefined comparison criteria are met; and

keeping the particular document as the representative document for the new set of documents if the set of predefined comparison criteria is not met.

48. (Original) The computer program product of claim 47, wherein the set of predefined comparison criteria comprise at least two parameters, one parameter for comparison with an absolute difference of document rank between the requesting document and the particular

document, and another parameter for comparison with a ratio of document rank between the requesting document and the particular document.

49. (Original) The computer program product of claim 40, wherein a document is a temporary redirect page comprising a document content, a source document address, and a target document address.

50. (Currently amended) A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for constructing a plurality of tables, each table corresponding to a portion of a document address space, storing information identifying documents having a same document content identifier and each identified document having an associated document rank; wherein documents having the same document content identifier have the same content and documents having different document content identifiers have different content;

instructions for receiving a newly crawled document, such document characterized by a document content identifier and a document rank;

instructions for reading information stored in the plurality of tables to identify a set of documents sharing the document content identifier of the newly crawled document, and ascertaining an original representative document for the identified set of documents;

instructions for updating the information stored in at least one of the tables in accordance with the document ranks of the identified set of documents and the newly crawled document;

instructions for determining a representative document for the newly crawled document and the identified set of documents;

instructions for indexing the representative document when said representative document is the newly crawled document; and

instructions for repeating the receiving, reading, updating, determining and indexing operations with respect to a plurality of newly crawled documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the newly crawled documents are determined to be representative documents and are indexed.

51. (Previously presented) The computer program product of claim 50, wherein information identifying the identified set of documents, including a particular document serving as the original representative document of the identified set, is stored in one or more tables.

52. (Previously presented) The computer program product of claim 51, wherein the determining includes

comparing the document rank of the newly crawled document with that of the particular document from the identified set in accordance with a set of predefined comparison criteria;

selecting the newly crawled document as the representative document if the set of predefined comparison criteria are met; and

keeping the particular document as the representative document if the set of predefined comparison criteria is not met.

53. (Previously presented) The computer program product of claim 51, wherein the set of predefined comparison criteria comprise at least two parameters, one parameter for comparison with an absolute difference of document ranks between the newly crawled document and the particular document, and another parameter for comparison with a ratio of document ranks between the newly crawled document and the particular document.

54. (Original) The computer program product of claim 50, wherein the updating includes inserting information identifying the newly crawled document into the at least one table only when a predefined insertion condition is satisfied.

55. (Previously presented) The computer program product of claim 50, wherein the predefined insertion condition is that the document rank of the newly crawled document is higher than the document rank of at least one document in the identified set of documents.

56. (Currently amended) A computer program product of detecting duplicate documents for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for constructing a plurality of tables, each table corresponding to a segment of a document address space, storing information identifying documents having a

same document content identifier and each identified document having an associated document rank, wherein the plurality of tables comprise N+1 tables where N is an integer greater than one, wherein the N+1 tables comprise N tables, each generated during a respective phase of a set of N crawling phases, and a current table generated during a current one of the N crawling phases, wherein an oldest one of the N tables was generated during a previous instance of the current crawling phase;

instructions for receiving a newly crawled document, such document characterized by a document content identifier and a document rank; wherein documents having the same document content identifier have the same content and documents having different document content identifiers have different content;

instructions for reading information stored in the N+1 tables to identify a set of documents sharing the document content identifier of the newly crawled document, and ascertaining an original representative document for the identified set of documents;

instructions for updating the information stored in the current table in accordance with the document rankings of the identified set of documents and the newly crawled document;

instructions for determining a representative document for the newly crawled document and the identified set of documents;

instructions for indexing the representative document when said representative document is the newly crawled document;

instructions for repeating the receiving, reading, updating, determining and indexing operations with respect to a plurality of newly crawled documents, each of which shares a respective document content identifier with a respective set of documents, such that at least some of the newly crawled documents are determined to be representative documents and are indexed; and

instructions for retiring the oldest one of the N tables upon completion of the current crawling phase.

57. (Original) The computer program product of claim 56, wherein the reading comprises reading from a merged table that stores information from a plurality of the N tables, and reading from the current table.

58. (Previously presented) The computer program product of claim 56, wherein the identified set of documents, including a particular document serving as the original representative document of the identified set, is stored in one or more tables.